UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

| APPLICATION NO. | FILING DATE | FIRST NAMED INVENTOR | ATTORNEY DOCKET NO. | CONFIRMATION NO. |
|---|---|---|---|---|
| 10/825,488 | 04/14/2004 | Venkatesh Ganti | 16-594 | 8552 |

22971 7590 12/05/2006

MICROSOFT CORPORATION
ATTN: PATENT GROUP DOCKETING DEPARTMENT
ONE MICROSOFT WAY
REDMOND, WA 98052-6399

| EXAMINER | |
|---|---|
| JOHNSON, JOHNESE T | |

| ART UNIT | PAPER NUMBER |
|---|---|
| 2169 | |

DATE MAILED: 12/05/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

| Office Action Summary | Application No. | Applicant(s) |
|---|---|---|
| | 10/825,488 | GANTI ET AL. |
| | Examiner | Art Unit |
| | Johnese Johnson | 2169 |

*-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

1)☒ Responsive to communication(s) filed on *14 April 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

4)☒ Claim(s) *1-34* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-34* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

**Application Papers**

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *14 April 2004* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *06 June 2005*.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

## DETAILED ACTION

### *Claim Objections*

1. Claims 1, 15, 17, 27 and 29 are objected to because of the following informalities:

Claim 1 recites: "for evaluating", which implies intended use.

Claim should be amended to recite "to evaluate".

Claims 1, and 27 recite: "for segmenting", which is intended use.

Claims should be amended to recite "to segment".

Claim 15 recites: "for performing", which is intended use".

Claim should be amended to recite "to perform".

Claim 17 recites: "for storing", which is intended use.

Claim should be amended to recite "to store".

2. Claim 13 is objected to because of the following informalities: claim is grammatically incorrect. The claim recites: "includes of states". The claim should be amended to recite "includes states".

3. Claim 29 is objected to because of the following informalities: "additionally including means for maintaining a collection of records is stored in a database relation". It is not clear to the examiner if the means is ***for storing*** a collection of records or if the actual means, itself ***is stored*** in a database relation. Appropriate correction is required.

### *Claim Rejections - 35 USC § 112*

4. The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5. Claim 1 is rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. The preamble recites, "for evaluating an input string". Evaluating an input string *is never* realized in the body of the claim; thus, there is no nexus between the intended use of the preamble and the body of the claim.

### *Claim Rejections - 35 USC § 101*

6. 35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 1, 16, 17, and 27 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1, 16 and 27 recite "determine[ing] a most probable segmentation of the input string" as the final step in the process. The act of *determining* does not produce any functional change, nor does it produce any useful, concrete, and tangible results. Therefore, these claims are non-statutory. Claims should be amended to produce a result/ output to the "determining" step i.e. storage or display.

Claims 1,16, 17, 22, and 27 are directed to program products. Program code is also known as functional descriptive material (See In re Warmerdam, 33 F3d at 1360,

31 USPQ2d at 1759). The content is not structurally and functionally interrelated to a computer-readable medium thereby rendering it incapable of producing a useful, concrete and tangible result and is therefore, non-statutory. The claims should be amended to recite hardware in the body of the claims.

## *Claim Rejections - 35 USC § 102*

7. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

8. Claims 1-13, 15, 17-19, 21, 22, 24-34 are rejected under 35 U.S.C. 102(b) as being anticipated by Borkar et al.; "Automatic segmentation of text strings into structured records".

As to claims 1 and 27, Borkar et al. disclose:

A process (see Abstract, pg. 1, line 1) and system (see Abstract, pg. 1, paragraph 2, line 1; wherein DATAMOLD is a system of interrelated components used to segment text) for evaluating an input string to segment said input string into component parts comprising:

means for providing a state transition model (see Abstract, pg. 1, paragraph 2, line 1 DATAMOLD) based on an existing collection of data records that includes probabilities for segmenting input strings into component parts which adjusts said

probabilities to account for erroneous token placement in the input string (see pg. 3, section 1.3.1, lines 19-21); and

means for determining a most probable segmentation (see Abstract, pg. 1, paragraph 2, line 1 DATAMOLD) of the input string by comparing an order of tokens that make up the input string with a state transition model derived from the collection of data records (see pg. 3, section 1.3.1, col. 2, lines 9-11; wherein the inner HMMs corroborate each other's findings to pick the segmentation that is globally optimal).

As to claims 2 and 28, <u>Borkar et al.</u> disclose:

wherein the state transition model has probabilities for multiple states of said model and a most probable segmentation is determined based on a most probable token emission path through different states of the state transition model from a beginning state to an end state (see pg. 4, col. 1, line 3; wherein the HMM has multiple states **and** col. 2, lines 6-9 –path having the highest probability).

As to claim 3, <u>Borkar et al.</u> disclose:

wherein the collection of data records is stored in a database relation and an order of attributes for the database relation as the most probable segmentation is determined (see pg. 3, Fig. 1; wherein the structured record is determined and produced).

As to claims 4 and 30, Borkar et al. disclose:

wherein the input string is segmented into sub-components which correspond to attributes of the database relation (see pg. 1, col. 2, section 1.1, lines 5-18).

As to claims 5 and 31, Borkar et al. disclose:

wherein the tokens are substrings of said input string (see pg. 6, section 2.4, lines 2-4).

As to claims 6 and 32, Borkar et al. disclose:

wherein the input string is to be segmented into database attributes and wherein each attribute has a state transition model based on the contents of the database relation (see pg. 4, Fig. 2; wherein each attribute has a transition in the model).

As to claims 7 and 33, Borkar et al. disclose:

wherein the state transition model has multiple states for a beginning, middle and trailing position within an input string (see pg. 6, Fig. 6; wherein state "1" is the beginning, state "2" is the middle and state "3" is the trailing position).

As to claims 8 and 34, Borkar et al. disclose:

wherein the state transition model has probabilities for the states and a most probable segmentation is determined based on a most probable token emission [state] path through different states of the state transition model from a beginning state to an end state (see pg. 6, Fig 6 and col. 2, paragraph 2, lines 1-4).

As to claim 9, Borkar et al. disclose:

wherein input attribute order for records to be segmented is known in advance of segmentation of an input string (see Abstract, pg. 1, paragraph 2, lines 3-8).

As to claim 10, Borkar et al. disclose:

wherein an attribute order is learned from a batch of records that are inserted into the table (see Abstract, pg. 1, paragraph 2, lines 1-3).

As to claim 11, Borkar et al. disclose:

wherein the state transition model has at least some states corresponding to base tokens occurring in the reference relation (see Abstract, pg. 1, paragraph 2, lines 1-8; wherein the training examples and dictionary provide the basis for acceptable and recognizable input and therefore some states would correspond to the same structure/ examples or base tokens).

As to claim 12, Borkar et al. disclose:

wherein the state transition model has class states corresponding to token patterns within said reference relation (see pg. 3, col. 1, paragraph 3, lines 1-8).

As to claim 13, Borkar et al. disclose:

wherein the state transition model includes of states that account for missing, misordered and inserted tokens within an attribute (see pgs. 3-4, section 2; wherein

data mold uses the example segmented records to output a model that when presented with any unseen text segments it into one or more of its constituent elements).

As to claim 15, Borkar et al. disclose:

A machine computer readable medium containing instructions for performing the evaluat[ion] [of] an input string to segment said input string into component parts (see pg. 1, section 1.1, lines 5-6; wherein the tool is used during warehouse construction which implies that the program instructions are being read from a medium inserted in or stored on a machine).

As to claim 17, Borkar et al. disclose:

A system for processing input strings to segment those records for inclusion into a database comprising:

a) a database management system for storing records organized into relations wherein data records within a relation are organized into a number of attributes (see page 1, Abstract, line 7 – corporate database);

b) a model building component that builds a number of attribute recognition models based on an existing relation of data records, wherein one or more of said attribute recognition models includes probabilities for segmenting input strings into component arts which adjusts said probabilities to account for erroneous entries within an input string (see page 1, Abstract, lines 13-14; wherein DATAMOLD comprises a model building component because its built on HMM); and

c) a segmenting component that receives an input string and determines a most probable record segmentation by evaluating transition probabilities of states within the attribute recognition models built by the model building component (see page 2, section 1.3, lines 1-3; wherein DATAMOLD comprises a segmenting component).

As to claim 18, Borkar et al. disclose:

wherein the segmenting component receives a batch of evaluation strings and determines an attribute order of strings in said batch and thereafter assumes the input string has tokens in the same attribute order as the evaluation strings (see Abstract, pg. 1, paragraph 2, lines 3-8; wherein the training examples are the batch of strings that provide a basis for the structure of strings).

As to claim 19, Borkar et al. disclose:

wherein the segmenting component evaluates the tokens in an order in which they are contained in the input string and considers state transitions from multiple attribute recognition models to find a maximum probability for the state of a token to provide a maximum probability for each token in said input string (see pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probablility).

As to claim 21, Borkar et al. disclose:

wherein the model building component defines a start and end state for each model and accommodates missing attributes by assigning a probability for a transition from the start to the end state (see pg. 6, Fig. 6).

As to claim 22, Borkar et al. disclose:

A string segmentation schema comprising: a state transition model for a data attribute of a data record wherein the transition model assigns token probabilities to a beginning, middle and trailing state of the model that are transitioned to from a start state and terminate with an end state (see Page 6, Fig. 6; wherein the state transition model has states for attributes of the input record and the edges represent the probabilities to the first (beginning state), second (middle state), third (trailing state)).

As to claim 24 Borkar et al. disclose:

wherein the schema includes a state transition models for multiple attributes of a database record and one or more of said models provide a transition probability between the start state and the end state of each attribute recognition model to accommodate missing attributes within an input string (see pg. 4, figure 2; wherein the model includes states for each attribute in an input string from a database record and the edges provide the probabilities between start and end states).

As to claim 25, Borkar et al. disclose:

A process of segmenting a string input record into a sequence of attributes for inclusion into a database table comprising:

considering a first token in a string input record and determining a maximum state probability for said token based on state transition models for multiple data table attributes (see pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probablility);

considering in turn subsequent tokens in the string input record and determining maximum state probabilities for said subsequent tokens from a previous token state until all tokens are considered (see pg. 4, section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probablility); and

segmenting the string record by assigning the tokens of the string to attribute states of the state transition models corresponding to said maximum state probabilities (see pg. 4, Fig. 2, wherein the model displays attributes represented by states and section 2.1; wherein the segmenting component considers transitions from the multiple attribute states to find the maximum probablility .

As to claim 26, Borkar et al. disclose:

additionally comprising determining an attribute order for a batch of string input records and using the order to limit the possible state probabilities when evaluating tokens in an input string (see Abstract, pg. 1, paragraph 2, lines 1-3; wherein the structure and order] is learned from the training examples).

### *Claim Rejections - 35 USC § 103*

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10. Claims 14, 20, and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al.; "Automatic segmentation of text strings into structured records" and in view of Reed (U.S. Pat. No. 5, 095, 432).

As to claim 14, Borkar et al. does not explicitly disclose:

wherein the state transition model has a beginning, a middle and a trailing state topology and the process of accounting for misordered and inserted tokens is performed by copying states from one of said beginning, middle or trailing states into another of said beginning, middle or trailing states.

However, Reed discloses:

wherein the state transition model has a beginning, a middle and a trailing state topology and the process of accounting for misordered and inserted tokens is performed by copying states from one of said beginning, middle or trailing states into another of said beginning, middle or trailing states (see col. 5, lines 1).

It would have been obvious, at the time of the invention, having the teachings of Borkar et al. and Reed before him/her, to combine the steps as disclosed by Borkar et al. with the feature as disclosed by Reed to enable grammar developers to use the

familiar PSG formalism to compile their grammars into RVG for more efficient execution (see Reed, col. 2, lines 54-57).

As to claims 20 and 23, Borkar et al. disclose:

wherein the model building component assigns states for each attribute for a beginning, middle and trailing token position (see pg. 4, Fig. 2; wherein the states are assigned to each attribute and pg. 6, Fig. 6; wherein states are assigned for first (beginning state), second (middle state), third (trailing state))

However, Borkar et al. does not explicitly disclose:

wherein the model building component relaxes token acceptance by the model by copying states among said beginning, middle and trailing token positions.

Reed discloses:

wherein the model building component relaxes token acceptance by the model by copying states among said beginning, middle and trailing token positions (see col. 5, lines 1; wherein states in the transition model are copied).

It would have been obvious, at the time of the invention, having the teachings of Borkar et al. and Reed before him/her, to combine the steps as disclosed by Borkar et al. with the feature as disclosed by Reed to enable grammar developers to use the familiar PSG formalism to compile their grammars into RVG for more efficient execution (see Reed, col. 2, lines 54-57).

11. Claim 16 is rejected under 35 U.S.C. 103(a) as being unpatentable over Borkar et al.; "Automatic segmentation of text strings into structured records" and in view of Fairweather (U.S. PG. Pub. No. 2006/0235811).

As to claim 16, Borkar et al. disclose:

A process for segmenting strings into component parts comprising:

providing a reference table of string records that are segmented into multiple substrings corresponding to database attributes (see Abstract, p. 1, paragraph 2, lines 1-3);

breaking the input record into a sequence of tokens, and determining a most probable segmentation of the input record by comparing the tokens of the input record with state models derived for attributes from the reference table (see pg. 3, section 1.3.1, col. 2, lines 9-11; wherein the inner HMMs corroborate each other's findings to pick the segmentation that is globally optimal).

However, Borkar et al. does not explicitly disclose:

analyzing the substrings within an attribute to provide a state model that assumes a beginning, a middle and a trailing token topology for said attribute said topology including a null token for an empty attribute component;

Fairweather discloses:

analyzing the substrings within an attribute to provide a state model that assumes a beginning, a middle and a trailing token topology for said attribute said topology including a null token for an empty attribute component (see paragraph [0406], lines 8-9; wherein a the null pointer is returned is token is null);

It would have been obvious, at the time of the invention, having the teachings of Borkar et al. and Fairweather before him/her, to combine the steps as disclosed by Borkar et al. with the feature as disclosed by Fairweather to provide a system in which the content of the data itself actually determines the order of execution of statements in the mining language and automatically keeps track of the current state (see Fairweather, paragraph [0004], lines 7-10).

## *Conclusion*

10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Johnese Johnson whose telephone number is 571-270-1097. The examiner can normally be reached on 4/5/9.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Christian Chace can be reached on 571-272-4190. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

J.J.

15 November 2006

J.J.

CHRISTIAN CHACE
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100